

## **Intermodal synchrony – as a form of maternal responsiveness – is associated with language development**

Katharina J. Rohlfing, Iris Nomikou  
Bielefeld University, NRW, Germany

Research findings indicate that synchrony between events in two different modalities is a key concept in early social learning. Our pilot study with 14 mother–child dyads is the first to support the idea that synchrony between action and language as a form of responsive behaviour in mothers relates to later language acquisition in their children. We conducted a fine-grained coding of multimodal behaviour within the dyad during an everyday diapering activity when the children were three and six months old. When the children attained 24 months, their mothers completed language surveys; this data was then related to the dyadic measures in early interaction. We propose a ‘switching-roles’ model according to which it is important for three-month-olds to be exposed to multimodal input for a great deal of time, whereas for six-month-old infants, the mother should respond to the infant’s attention and provide multimodal input when her child is gazing at her.

**Keywords:** acoustic packaging, multimodal maternal responsiveness, early mother-infant interaction, social language learning

### **1. Introduction**

Language and action are commonly viewed as two different skills: Whereas the development of action understanding and production is considered to be a part of the motor system, language acquisition is linked to the development of social and cognitive skills. However, there are good reasons to claim that language and action work tightly together in child development. In the following, we will show that even in early interactions, language can already be perceived as a social signal segmenting the action stream meaningfully. This complements the positive effect of synchronous action presentation on the acquisition of linguistic items reported in a number of studies.

#### ***1.1 Language scaffolds action segmentation / action understanding***

Early in development, before semantic content can be processed, language already functions as a social signal. This signal has the power to be perceived as a marker. Newborns are

already sensitive to human speech (Vouloumanos & Werker 2007) and can attend to their mothers' voices even in the presence of noise (Bortfeld, Shaw & Depowski 2013). In addition, infants prefer human voices over other sounds (Colombo & Bundy 1983; Muir & Field 1979). This social signal does not just play a role in emotional attunement between the caregiver and the child (Markova & Legerstee 2006), it also modulates infants' attention to objects and events. A number of studies on infant categorization have shown that infants perceive visual presentation better when it is accompanied by speech (Horowitz 1974; McCartney & Panneton 2005). Moreover, using words – but not tones – to highlight commonalities or differences between objects facilitates the categorization of these objects by three-month-olds (Ferry, Hespos, & Waxman 2010; also Balaban & Waxman 1997; Plunkett, Hu & Cohen 2008). Thus, infants perceive language as a social signal with definite priority. In natural input to three-month-old children during everyday activities such as diapering, language seems to be present to a great degree when mothers interact with them (Nomikou & Rohlfing 2011). Language and actions seem to be highly coordinated in this multimodal input (*ibid*; Nomikou, Rohlfing & Szufnarowska 2013). Such tight audio–visual coordination was also observed in studies analysing the input to older children at the age of 8 to 12 months. Without analysing what is said in an interaction, Rolf, Hanheide and Rohlfing (2009; also Schillingmann, Wrede & Rohlfing 2009) have shown that the audio–visual coordination seems to be greater in children-directed interaction than in adult–adult-directed interaction. Schillingmann et al. (2009) analysed how far adults 'package' events when demonstrating the function of objects to their 9- to 12-month-old children. By packaging we mean a temporal overlap of speech and action signal. The authors found that more acoustic packages were presented in child-directed than in adult–adult-directed input. In addition, the former packages contained less content, because they contained fewer motion peaks. Focusing more on what is said during action demonstrations, Meyer, Hard, Brand, McGarvey and Baldwin (2011) have shown that mothers' speech is synchronous to their actions when interacting with 6- to 13-month-old infants. More specifically, when demonstrating how to perform some actions, mothers seem to align speech with related action (rather than aligning speech in general with actions), thereby providing a meaningful multimodal behaviour (see also Gogate, Bahrick & Watson 2000).

The fact that language is presented synchronously to actions also facilitates the development of infants' action understanding. To clarify this problem, Brand and Tapscott (2007:321) describe the following scenario: "An actor walks across the kitchen floor, grasps and releases a small towel on the counter beside the sink, turns the water on, touches a bar of soap, rubs her hands together under the water, turns the water off, then grasps and releases the towel again". If an observer has never seen such an event before, the question is how she or he might determine that the first grasping of the towel is included in the hand-washing event.

Clearly, a naïve observer, who is not familiar with the meaning of these actions, needs support in the action segmentation (Zukow-Goldring 2006). Recognizing these segments can lead to the emergence of a hierarchically organized structure in action (Zacks & Tversky 2001). Adults, spontaneously encode an event in its boundaries such as its goals, parts and sub-parts (Zacks & Swallow 2007). Hirsh-Pasek and Golinkoff (1996) proposed that providing overlapping speech to some crucial parts of an ongoing event will mark some event boundaries, and infants will perceive these marked units better. More specifically, Hirsh-Pasek and Golinkoff (1996) suggested that parts of actions that overlap temporally with speech will form acoustic packages. These packages provide meaning to children's action perception. Brand and Tapscott (2007) demonstrated how such packages might scaffold the process of action processing. In a preferential looking study, 9.5-month-old children were presented with a sequence of actions (A–B–C) in a familiarization phase. The actions were accompanied by language in the form of a narration such as “Wow! Do you see what she's doing? She's blixing!” These overlapped temporally with some actions (e.g. A–B or B–C). The authors found that in the test phase, infants considered that those sequences of actions that were ‘packaged’ by the narration belonged together.

What is the basis for these scaffolding effects of language on infant's action perception and processing? There are two important notions here: attention and synchrony. Selective attention is most pronounced especially in infants rather than in older children (Bahrick, Lickliter & Flom 2004). Synchrony refers to the fact that events in two modalities (such as hearable sound and visible movement) can be coordinated with each other temporally so that there appears to be a correlation between the characteristics of two events (Cangelosi et al. 2010). However, the degree to which the two events can overlap varies (see Nomikou & Rohlfsing 2011). In the framework of the intersensory redundancy hypothesis (Bahrick et al. 2004), the basic idea is that stimulus properties that are synchronous, and thus redundant, reinforce each other. On this basis, they become foreground and attract the infants' attention. According to this mechanism, processing priority in early development is given to redundantly specified stimulus properties in contrast to other properties that become background. “Thus, the sights and sounds of hands clapping provide intersensory redundancy because they are temporally synchronous, originate in the same place, and convey the same rhythm, tempo, and intensity patterns in vision and audition” (Bahrick et al. 2004:100).

To summarize, we have argued that primitive meanings in event perception are formed when scaffolded with language input. On this early developmental level, infants are sensitive to language and can use it to guide their attention to redundantly specified parts of an ongoing event. Below, we will show not only that language can scaffold action perception but also that action can enrich language processing.

### ***1.2 Synchronous action contributes to the reference of language***

Naturalistic and experimental research has demonstrated that infants learn presented verbal behaviour better when it is coupled with synchronous actions than when it is presented asynchronously to action. Zukow-Goldring (1996), for example, investigated mother–child dyads in naturalistic settings and observed that mothers help their children perceive meaningful language by providing “additional perceptual structure” (Zukow-Goldring 1996:195). What she means by “perceptual structure” is not only different forms of gestures but also a tight temporal coupling of performed actions with language. In a more limited setting in which the participating mothers were instructed explicitly to teach new labels for objects and events to their children, mothers of early lexical children (5 to 17 months old) were found to move the new objects in a synchronous way when providing the new label (Gogate et al., 2000). In further controlled experiments six-month-old (Gogate & Bahrick 2001) and even two-month-old infants (Gogate, Prince & Matatyaho 2009) heard a syllable and were presented with an object that was moved either synchronously or asynchronously to the syllable. Gogate and her colleagues hypothesized that the synchronous movement of the referent when hearing a syllable would facilitate lexical mapping in infants. In line with this hypothesis, they found that infants remembered the syllable–object relations and could detect changes in novel syllable–object pairing only when the referent moved synchronously in time to the syllable provided during learning.

In addition, using eye-tracking technology, de Villiers Rader and Zukow-Goldring (2010) presented videos to 9- to 15-month-old infants in which a person was showing a new object and labelling it in either a (a) synchronous, (b) asynchronous or (c) static gesture condition. In the synchronous condition, the actor moved the object forward (looming motion) at the time at which it was labelled. In the asynchronous condition, the movement of the object began as soon as the actor started the script of labelling the object. Thus, in this condition, a movement was visible, but it was not synchronized temporally with the label of the object. Finally, in the static gesture condition, the actor held the object and did not move it at all. The authors found that 9- to 15-month-old infants did not just prefer looking at objects that were presented in a synchronous word–object condition; in this condition, they also demonstrated a better comprehension of the word. Whereas the synchrony as a form of “perceptual structure” (Zukow-Goldring 1996:195) seems to be important for young children up to the age of 15 months, between 15 and 19 months, children demonstrate the use of other social (rather than perceptual) cues to develop the meaning of a referent (Gogate et al. 2000; Golinkoff & Hirsh-Pasek 2006), and the role of temporal synchrony fades (Gogate et al. 2000).

It seems that providing synchrony between action and language is a form of tutoring behaviour that occurs when addressing a child. However, some studies reveal that this

tutoring behaviour is highly responsive to the learner's attention and her or his skills. For example, a study by Fischer, Foth, Rohlfing and Wrede (2011) using an artificial learner (a robot) showed that tutors provide a modified behaviour only in those modalities that are used as feedback by the learner. More specifically, in this study, the robot provided only visual feedback in the form of eye movements. Correspondingly, tutors modified their behaviour in action but not in their speech performance.

In a recent naturalistic and longitudinal study, Nomikou and colleagues (2013) analysed how far mothers provide input synchronous to the child's attention. They defined attended synchrony as the period of maternal behaviour that overlapped with the child's eye gaze towards the mother. The authors found that the infant's attention triggered the multimodal input in the sense that "episodes of eye contact were reinforced systematically by accompanying verbal and nonverbal behaviors" (ibid:261).

In sum, the studies presented above show, on the one hand, that by synchronizing speech and action, caregivers recruit infants' attention and create a non-arbitrary link between word and world (Zukow-Goldring 1996). On the other hand, caregivers are responsive and attuned to children's initiatives, and they adapt their behaviour in developmentally appropriate ways (Gogate et al. 2000; Tamis-LeMonda, Bornstein & Baumwell 2013). Thus, the (attended) synchrony between two modal events appears to be a key concept in early social learning (Cangelosi et al. 2010; Rolf et al. 2009).

However, there is still little known about whether synchrony between action and language as a form of responsive behaviour predicts infants' language skills. Up to now, analyses of maternal responsiveness studies have considered only related skills. More specifically, verbal responsiveness has been found to correlate positively with a variety of language skills including the acquisition of lexicon and grammar (e.g. Masur, Flynn & Eichorst 2005; Tamis-LeMonda, Bornstein & Baumwell 2001). However, as has been recently suggested, maternal responsiveness is a multi-dimensional construct (McGillion et al. 2013) including not only verbal but also semantically and temporally contingent responses. Interestingly, in this recent study, the authors found that utterances that were both semantically appropriate and temporally linked to an infant vocalization at 9.5 months related to infant expressive vocabulary at 18 months (McGillion et al. 2013). Clearly, the temporal dimension seems to be crucial in younger infants.

The study presented here was motivated by the mutual influence of language and action. We reasoned that providing language as a social signal accompanying actions and events early in infancy might facilitate infants' action processing. A more elaborated understanding of action, in turn, might provide a robust conceptual basis for the meaning of action words. Following this logic, we hypothesized that the variability in a mother's responsive synchrony in early interactions with her child will correlate with the variability of the child's language

acquisition (especially the acquisition of verbs that denote events). Our logic is consistent with the embodied and distributed view of language (Cowley 2011; Rączaszek-Leonardi 2009) according to which early language development progresses from coordination of multimodal behaviour on a temporal level to joint actions in the form of communicative and meaningful acts (Rączaszek-Leonardi, Nomikou & Rohlfing 2013).

## **2. Method**

Our study investigated the question whether the temporal synchrony between action and language as a form of caregivers' responsiveness in early interaction behaviour shapes children's language learning. For our purpose, we observed natural interactions between mothers and infants when the infants were three and six months old. We focussed on the changing capacities of the infants' gaze behaviour and the multimodal behaviour of the mother in interactions. In previous work, we have reported that mothers vocalize in a tight relationship with action, 'packaging' their actions acoustically and making language perceivable and tangible to the infants (Nomikou & Rohlfing 2011). We suggested that the congruent overlap of language and action assist infants in discovering that sounds relate to or refer in some way to parts of events and objects, thus creating a first form of reference. However, we found that caregivers use significantly more synchrony when acting within the focus of their infant's gaze, and that they also sustain the synchrony for the duration of its overlap with their infant's gaze (Nomikou et al. 2013). Thus, synchrony (whether perceived by the child or not) was found to be crucial (ibid).

Following on from these studies, we observed a great variability between subjects in the extent to which (1) maternal synchrony is responsive to the infant's gaze, and (2) infants attend to the synchrony provided by the mother. These findings raised the question whether this variability could predict the infant's language development.

In the follow-up study presented here, we collected data with the ELFRA-2 (Grimm & Doil 2000), a parental questionnaire (German adaptation of the Infant Form of the MacArthur-Bates Communicative Development Inventories) at the age of 24 months in order to assess the infants' language development. We were interested in the overall productive vocabulary and the proportion of spoken verbs. Our hypothesis was that the packaging of actions with vocal behaviour would have an effect on the infants' later vocabulary. In particular, we predicted that because acoustic packaging is associated with event processing in younger infants (Brand & Tapscott 2007), we should obtain a correlation between the maternal behaviour and verbs as words for events.

### **2.1 Participants**

Fourteen mothers and their children (nine boys and five girls) were recruited for the

longitudinal study in the city of Bielefeld and its surroundings. Ten of the children were firstborn. The average age of all participating children was 15 weeks ( $SD = 1$ ) at the first visit and 27 weeks ( $SD = 1$ ) at the second visit. All infants had been born with no complications. The average age of the mothers was 32 years ( $SD = 3.8$ ); they had all completed secondary education and were native speakers of German.

## **2.2 Procedure**

To record the mother–child interactions, participants were visited in their homes. To provide similar conditions for the actions performed during the diapering, a foldable changing table was brought to the home. This was placed in the centre of a spacious room (either the living or the child’s room). Two HD video cameras were positioned around the table: one focusing on the mother’s upper body and the other focusing on the child and her or his body (see Figure 1). An external microphone was mounted on one of the cameras to guarantee high-quality audio recording. Further details of the setting are described in Nomikou et al. (2013).

-----  
@Insert Figure 1: Camera setup  
-----

For the parental language survey (ELFRA-2), mothers were contacted when their children were 24 months old. They received instructions on how to fill out the survey by phone. The survey was then sent to the participants via mail. Their written reports were sent back after several days. The mothers were experienced with this type of measurement because they had already filled out ELFRA-1 when their children were 12 months old. ELFRA-1, however, has been criticized for not being sensitive to language development aspects (Sachse, Saracino & von Sucholodetz 2007). Therefore, we do not report this data here. In our analysis of reported verb production, we focussed on two sections of the checklist including 14 modal verbs types and 31 event verbs types. The majority of the event verbs were dynamic. We reimbursed mothers for their effort with a book for their child.

## **2.3 Coding**

We processed the data on language–action synchrony by coding video recordings with INTERACT 9, an observational data transcription software. Two separate coders used frame-to-frame analysis to mark the onset and offset of minimal segments during which the mothers’ body movement (head, hands, torso and facial expressions) and her verbal behaviour co-occurred. These multimodal intervals produced our basic units of analysis, which we call *maternal synchrony*. This is a behaviour in which maternal language is accompanied by

bodily movements (Nomikou & Rohlfsing 2011) as can be seen in Figure 2.

-----  
@Insert Figure 2  
-----

Furthermore, we coded gaze direction (gaze at the interaction partner, gaze at object or interaction partner's body parts and gaze away) for both mother and infant. We calculated the following measures from the overlap between infant's gaze towards the mother and the mother's synchrony intervals:

*Infant's attention to synchrony*, representing the proportion of the infant's gaze towards the mother overlapping with maternal synchrony (s. Figure 3) (i.e. the duration in seconds of the overlap of infant gaze with maternal synchrony was divided by the total duration in seconds of infant gaze at the mother)

*Maternal synchrony* is the overall duration in seconds of synchrony events (i.e. the maternal behaviour in which language was emphasized through bodily movement) divided by the length of the video in seconds;

*Attended synchrony* (see Figure 3), representing the proportion of maternal synchrony overlapping with the infant's gaze towards the mother. (i.e. the duration in seconds of overlap of maternal synchrony with infant gaze was divided by the total duration in seconds of maternal synchrony).

To control for the possibility that all mentioned measures depend on mother's verbal behavior, we also coded: *Maternal talkativeness*, i.e. the duration of maternal verbal behavior as a proportion over the interaction time.

-----  
@Insert Figure 3  
-----

A naïve coder (with no coding experience with the particular data) coded 13% of the data for maternal synchrony and 11% for gaze. Intercoder agreement was then calculated and ranged between 89% and 100%; *Cohen's kappa* = 0.8 for maternal synchrony and between 97%–100%; *Cohen's kappa* = 0.9 for maternal gaze. For infant gaze, agreement ranged between 86%–89%; *Cohen's kappa* = 0.6.

### 3. Results

Table 1 presents all the data for the individual subjects as assessed by the variables presented above. The table also presents the children's reported productive vocabulary (ELFRA-2 prod.) as well as the percentage of verbs in the overall reported vocabulary (ELFRA-2 verbs



%).<sup>1</sup> At the bottom of the table, the data are summarized in the form of descriptive statistics (minimum value, maximum value, mean and standard deviation).

-----  
@Insert Table 1  
-----

We observed a great deal of variability between subjects in the extent to which (1) maternal synchrony was responsive to the infant’s gaze and (2) infants attended to the synchrony provided by the mother. There was also (3) a variation in the child’s reported vocabulary. We therefore conducted correlational analyses between the dependent measures to assess the relationships (see Table 2).

-----  
@Insert Table 2  
-----

**3.1 Early interaction at three months of age and child’s language learning**

We found that infants’ attention to synchrony at the age of three months correlated significantly with the overall reported productive vocabulary ( $r = .525, p < .05$ ) and the proportion of reported spoken verbs ( $r = .530, p < .05$ ) at the age of 24 months. This finding suggests a relation between the amount of synchrony that the children attend to in early interaction and their later vocabulary skills: The more mothers provided synchrony between their action and language when their infants were three months old, the better the infants’ reported overall productive vocabulary and ability to produce verbs at the age of 24 months.

**3.2 Early interaction at six months of age and child’s language learning**

The roles of the mother and the child seemed to change as the child develops. When the children were six months old, just providing a synchronized action and language behaviour was no longer sufficient, as shown by the lack of a significant correlation between the amount of maternal synchrony and children’s reported verb production or overall reported production at 24 months (see Table 2). Instead, we found a significant correlation between maternal attended synchrony and productive vocabulary ( $r = .619, p < .05$ ) as well as the overall proportion of reported spoken verbs ( $r = .552, p < .05$ ).

---

<sup>1</sup> To assess the language performance in children, we also conducted a standardized test – SETK-2 (Grimm 2000) – in a laboratory setting when the children were 24 months old. However, the sample of children participating in the SETK-2 testing ( $N = 9$ ) was only a subpart of the sample from whom we received an ELFRA ( $N = 14$ ). We computed correlations between these two measurements, the reported vocabulary and the experimentally tested language performance: Kendall’s tau  $\tau = .817, n = 9, p = .002$  for language production and Kendall’s tau  $\tau = .535, n = 9, p = .046$  for verb production. Due to limitations of space, this article presents only the data from the language survey.

### 3.3 'Switching the roles'-model

The two variables, infants' attention to synchrony at three months and mother's responsive synchrony when the infants were six months old, were not independent. The correlation between the two was highly significant ( $r = .80, p < .001$ ), suggesting that the maternal behaviour when the child was three months old related strongly to her behaviour when the child was aged six months.

A stepwise regression analysis assesses the independent contribution of each of the predictor variables on the dependent variable (ELFRA-2). When, as in our case, the two predictors correlate with each other, this raises the problem of multicollinearity because neither of the input variables appears to be a significant predictor. Therefore, we calculated a linear regression to gain insights into how these two variables related to later language acquisition skills. The analysis showed that when considering independently from each other, infants' attention to synchrony at three months explained 28% of the ELFRA-2 variance ( $F(1,12) = 4.56, p < 0.05$ ) and attended synchrony at six months 38% ( $F(1,12) = 7.45, p = 0.05$ ) of this variance. In addition, both analyses revealed large effect sizes. Thus both variables were predictive of the overall productive vocabulary at 24 months. With respect to the proportion of spoken verbs at 24 months, both variables, maternal synchrony at three months (28%) and attended synchrony at six months (30%), significantly predicted the variance of the lexicon. This analysis confirmed our observation that the amount of maternal temporal coordination of action with speech perceived by the infant in early interaction when the child is three months old, and the extent to which mothers provide this temporal synchronization in accordance with child's attention later when the child is six months old seem to support later language development (see Figure 4).

-----  
@Insert Figure 4  
-----

One could argue that not the synchrony as a form of responsiveness but rather the verbal behaviour is related positively to the child's vocabulary development. Table 1 shows that one subject (Subject 13), who seemed to be delayed in ELFRA-2 values, was exposed to very little verbal interaction. To investigate this possibility, we calculated the talkativeness of the mothers as a proportion of the interaction time that they devoted to verbal activities. We found that at both ages talkativeness of the mother did not relate to the child's reported vocabulary (see Table 2). This finding is supportive of our model proposing that a multimodal behavior rather than only verbal input is predictive of language development.

### ***3.4 Reported lexicon***

To gain a more detailed picture of the child's reported vocabulary skills, we conducted analyses of different variables from the reported lexicon. We found a highly significant correlation between the proportion of spoken verbs and the overall productive vocabulary ( $r = .926, p < .001$ ). This suggests that the more verbs the children were reported to produce, the more extended was their overall vocabulary. This is consistent with the view that verb learning is fundamental to language development (Bates et al. 1994; Göksun, Hirsh-Pasek & Golinkoff 2010). Among the verbs, the modal verbs correlated significantly with the event verbs ( $r = .84, p < .001$ ). In our analyses, we therefore considered both types of verbs together.

### ***3.5 Applying the switching-roles model to children with low vocabulary: a case study***

Our analysis of individual differences in language acquisition revealed that for two children (Subjects 09 and 13 in Table 1), the lowest values for productive vocabulary in our sample were reported. However, because the mother of Subject 13 did not speak during the interaction with her six-month-old child, we will describe only Subject 09 in more detail. For our fine-grained analysis, we conducted a median split of the whole sample to obtain a critical value for maternal synchrony at a child's age of three months ( $Md = 36.67$ ) and maternal responsive synchrony at a child's age of six months ( $Md = 42.01$ ). We noticed that both measures were below the threshold for Subject 09 (see Table 1). This finding provides further support for the relation between maternal behaviour in early interaction and subsequent productive language development.

## **4. Discussion**

Drawing on findings in the literature reporting that synchrony between events in two modalities appears to be a key concept in early social learning (Bahrick et al. 2004; Cangelosi et al. 2010), our pilot study aimed to shed more light on responsiveness as a multimodal phenomenon. Up to now, little is known about how the synchrony between action and language as a form of responsive behaviour (Gogate et al. 2000; Nomikou & Rohlfing 2011; Nomikou et al. 2013; Zukow-Goldring 1996) relates to later language acquisition. In this study, we argued that providing language as a social signal accompanying actions and events early in infancy might facilitate an infant's action processing. In turn, a more elaborate understanding of action might provide a robust conceptual basis for the meaning of action words. Following this logic, we hypothesized that variability in the infant's attention to the mother and in the mother's responsiveness in early interactions would correlate with

variability in the child's language acquisition and, especially, with variability in the acquisition of verbs.

We conducted an extensive, fine-grained coding of maternal verbal and action behaviour as well as gaze behaviour during a natural everyday diapering activity in 14 mother–infant dyads when the children were three and six months old (Nomikou & Rohlfing 2011; Nomikou et al. 2013). When the children were 24 months old, we asked their mothers to complete a language survey.

For the 3 month-olds, we found a positive and significant correlation between the amount of synchrony the children attended to in early interaction and their later vocabulary skills. It should be noted that during early interactions at this young age, the child is looking at the mother to a greater extent (Nomikou et al. 2013). Hence, interaction has been suggested as being in the “affective reciprocal phase” (Colas 1999:114) at the age of three months. At this stage, interactions are dyadic and interpersonal, meaning that the main interest lies in coordinating attention within the dyad (Schaffer 1984). Thus, because the child focuses on maternal behaviour anyway, the positive relation to later language acquisition seems to be fostered by the mother providing synchronized action and language behaviour at this age. In other words, at this young age of three months, mothers' behaviour in early interaction seems to already foster the child's later language development by providing synchrony between actions and speech.

When the infants were six months old, just providing synchronized action and language behaviour no longer seemed to suffice. Instead, we found a significant correlation between maternal attended synchrony and children's later reported vocabulary. This finding strongly suggests that at the older age, it becomes important for the mother to monitor the focus of attention of her child who is now no longer just looking at her but is also attracted and distracted by events and objects in the surroundings (Bakeman & Adamson 1984). In this vein, Nomikou et al. (2013:258) observed that “mothers and the infants ‘switch’ their gazing roles between the ages of 3 and 6 months”. Thus, while at the age of three months, the child focuses on the mother and attends to her responsive behaviour anyway, at the age of six months, it is the mother who needs to actively search for slots of infant attention to her and to make her responsiveness perceivable. In other words, our explorations suggest that to foster the child's language learning when the child is six months old, the mother needs to provide synchrony between her speech and action precisely when the child is attending to her.

Incorporating both the behaviour of the mother and the infant during early interaction into two dyadic measures, we investigated the predictive role of the temporal coordination between language and action as a form of maternal responsiveness to infant attention. On the basis of our results on the whole sample as well as of individual children potentially at risk for language delay, we developed a ‘switching-the-roles’ model, according to which three-

month-old infants spend time looking at the mother during interaction. At six months of age, however, it is the mother who is responsive to the infant's attention. Our results revealed that three-month-old infants who perceived more synchrony provided by their mothers produced more verbs and had larger vocabularies overall at 24 months. Mothers who were more responsive to the gaze of their six-month-old infants when providing their synchrony had toddlers who had overall larger vocabularies at 24 months and produced more verbs than their peers. Verb production might, however, reflect the critical vocabulary size that was overall reached by the children rather than their advantage in discerning a particular grammatical structure (Bates et al. 1994). Further research needs to address experimentally the possibility that the acquisition of verbs may support an infant's perception of semantic relationships between objects and participants in events, thus leading to a broader vocabulary altogether.

It might be surprising to read that our analyses did not reveal a relationship between maternal talkativeness during early interaction and the children's later vocabulary development. It is important to note, however, that our measure for talkativeness encompasses only the quantity of speech. In the literature, in contrast, specific forms of maternal verbal behaviour (e.g. semantically contingent talk) were extensively discussed as facilitating vocabulary growth rather than the amount of speech overall (s. McGillion et al. 2013 for a recent overview). Our results thus add to this line of research suggesting that instead of considering only the quantity of provided speech, it is important to regard verbal behaviour in the context of interaction and (the age of) interlocutors.

We are aware of the limitations of our longitudinal study, which was observational in its nature. Because we cannot experimentally control for the precise effects of the two variables, infants' attention to synchrony and mothers responding with synchrony to the gaze of their infant, on language development at 2 years, the model should be considered as a hypothesis for future experimental studies. Accordingly, even in early interaction, language may already function as a social signal packaging and structuring events for infants' perception. Mothers who are responsive to their infants' engagement and provide multimodal language accompanying sensory rich dynamic events may make the meaningful action units more salient. This seems to support the acquisition of vocabulary.

Our pilot study is the first to indicate that maternal responsiveness comprises different aspects of the multimodal input at different ages: At three months of age, it is important for the infant to be exposed to multimodal input for a great deal of time, whereas at six months, it seems important for the input to be responsive to the infant's attention. In this sense, being responsive and communicating with bodily movement rather than via semantics might tune to and promote later language skills: "[M]oving with others" can tune to and promote a later "speaking with others" (Rączaszek-Leonardi et al. 2013). This is in line with an embodied and distributed view of language (Cowley 2011; Rączaszek-Leonardi 2009), according to

which early language development progresses from coordination of multimodal behaviour on a temporal level to joint actions in form of communicative and meaningful acts (Rączaszek-Leonardi et al. 2013). When learning, children take advantage of interaction that spreads information across learning domains. Further research needs to verify our observational findings by applying experimental designs to study the interdependence between particular multimodal interaction styles and children's vocabulary.

Finally, our observations contribute to the field of language acquisition research by suggesting that when working with longitudinal data, it is necessary to consider that the measures used for comparing different ages are age-dependent, because different factors or cues might be weighted differently at various stages of development (see also Hollich, Hirsh-Pasek & Golinkoff 2000).

## REFERENCES

- Bahrick, L.E., Lickliter, R. & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception and cognition in infancy. *Current Directions in Psychological Science* 13, 99–102.
- Bakeman, R. & Adamson, L.B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development* 55, 1278–1289.
- Balaban, M.T. & Waxman, S.R. (1997). Do words facilitate object categorization in nine-month-old infants? *Journal of Experimental Child Psychology* 64, 3–26.
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language* 21, 85–123.
- Bortfeld, H., Shaw, K. & Depowski, N. (2013). Disentangling the influence of salience and familiarity on infant word learning: methodological advances. *Frontiers in Psychology* 4, 174.
- Brand, R.J. & Tapscott, S. (2007). Acoustic packaging of action sequences by infants. *Infancy* 11, 321–332.
- Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K. Tani, J., Belphaeme, T., Sandini, G., Fadiga, L., Wrede, B., Rohlfing, K., Tuci, E., Dautenhahn, K., Saunders, J. & Zeschel, A. (2010). Integration of action and language knowledge: a roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development* 2, 167–195.
- Colas, A. (1999). Introducing infants to referential events: a development study of maternal ostensive marking in French. *Journal of Child Language* 26, 113–131.
- Colombo, J. & Bundy, R. (1983). Infant response to auditory familiarity and novelty. *Infant Behavior and Development* 6, 305–311.
- Cowley, S. J. (2011). Taking a language stance. *Ecological Psychology* 23, 185–209.
- Ferry, A.L., Hespos, S.J. & Waxman, S.R. (2010). Categorization in 3- and 4-month-old infants: an advantage of words over tones. *Child Development* 81, 472–479.
- Fischer, K., Foth, K., Rohlfing, K.J. & Wrede, B. (2011). Mindful tutors: linguistic choice and action demonstration in speech to infants and robots. *Interaction Studies* 12, 134–161.
- Gogate, L.J., & Bahrick, L.E. (2001). Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable-object relations. *Infancy* 2, 219–231.
- Gogate, L.J., Bahrick, L.E. & Watson, J. (2000). A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child Development* 71, 878–894.

- Gogate, L.J., Prince, C.G. & Matatyaho, D.J. (2009). Two-month-old infants' sensitivity to changes in arbitrary syllable-object pairings: the role of temporal synchrony. *Journal of Experimental Psychology: Human Perception and Performance* 35, 508.
- Göksun, T., Hirsh-Pasek, K. & Golinkoff, R.M. (2010). Trading spaces: carving up events for learning language. *Perspectives on Psychological Science* 5, 33–42.
- Golinkoff, R.M. & Hirsh-Pasek, K. (2006). The emergentist coalition model of word learning in children has implications for language in aging. In E. Bialystok & F. Craik (eds.), *Lifespan cognition: mechanisms of change* (207–222). New York, NY: Oxford University Press.
- Grimm, H. (2000). *Sprachentwicklungstest für zweijährige Kinder (SETK-2)*. Göttingen: Hogrefe.
- Grimm, H. & Doil, H. (2000). *Elternfragebögen für die Früherkennung von Risikokindern (ELFRA)*. Göttingen: Hogrefe.
- Hirsh-Pasek, K. & Golinkoff, R. M. (1996). *The origins of grammar*. Cambridge, MA: MIT Press.
- Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. (2000). The emergentist coalition model. *Monographs of the Society for Research in Child Development* 65, 17–29.
- Horowitz, F.D. (1974). *Visual attention, auditory stimulation, and language discrimination in young infants*. Chicago: University of Chicago Press (*Monographs of the Society for Research in Child Development* Nr. 39).
- Markova, G. & Legerstee, M. (2006). Contingency, imitation, and affect sharing: Foundations of infants' social awareness. *Developmental Psychology* 42, 132–141.
- Masur, E.F., Flynn, V. & Eichorst, D.L. (2005). Maternal responsive and directive behaviours and utterances as predictors of children's lexical development. *Journal of Child Language* 32, 63–91.
- McCartney, J.S. & Panneton, R. (2005). Four-month-olds' discrimination of voice changes in multimodal displays as a function of discrimination protocol. *Infancy* 7, 163–182.
- McGillion, M.L., Herbert, J.S., Pine, J.M., Keren-Portnoy, T., Vihman, M.M. & Matthews, D.E. (2013). Supporting early vocabulary development: What sort of responsiveness matters? *IEEE Transactions on Autonomous Mental Development* 5(3), 240–248.
- Meyer, M., Hard, B., Brand, R.J., McGarvey, M. & Baldwin, D.A. (2011). Acoustic packaging: Maternal speech and action synchrony. *IEEE Transactions on Autonomous Mental Development* 3, 154–162.
- Muir, D. & Field, J. (1979). Newborn infants orient to sound. *Child Development* 50, 431–436.
- Nomikou, I. & Rohlfing, K.J. (2011). Language does something. Body action and language in maternal input to three-month-olds. *IEEE Transactions on*



- Autonomous Mental Development* 3, 113–128.
- Nomikou, I., Rohlfing, K.J. & Szufnarowska, J. (2013). Educating attention: Recruiting, maintaining and framing eye contact in early natural mother-infant interactions. *Interaction Studies* 14, 240–267.
- Plunkett, K., Hu, J.-F. & Cohen, L.B. (2008). Labels can override perceptual categories in early infancy. *Cognition* 106, 665–681.
- Rączaszek-Leonardi, J. (2009). Symbols as constraints. The structuring role of dynamics and self-organization in natural language. *Pragmatics & Cognition* 17, 653–676.
- Rączaszek-Leonardi, J., Nomikou, I. & Rohlfing, K.J. (2013). Young children's dialogical actions: The beginnings of purposeful intersubjectivity. *IEEE Transactions on Autonomous Mental Development* 5(3), 210–221.
- Rolf, M., Hanheide, M. & Rohlfing, K.J. (2009). Attention via synchrony. Making use of multimodal cues in social learning. *IEEE Transactions on Autonomous Mental Development* 1, 55–67.
- Sachse, S., Saracino, M. & v. Sucholodetz, W. (2007). Prognostische Validität des ELFRA-1 bei der Früherkennung von Sprachentwicklungsstörungen. *Klinische Pädiatrie* 219, 17–22.
- Schaffer, H.R. (1984). *The child's entry into a social world*. NY: Academic Press
- Schillingmann, L., Wrede, B. & Rohlfing, K.J. (2009). A computational model of acoustic packaging. *IEEE Transactions on Autonomous Mental Development* 1, 226–237.
- Tamis-LeMonda, C.S., Bornstein, M.H. & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development* 72, 748–767.
- Tamis-LeMonda, C., Kuchirko, Y. & Tafuro, L. (2013). From action to interaction: Infant object exploration and mothers' contingent responsiveness. *IEEE Transactions on Autonomous Mental Development* 5(3), 202–209.
- de Villiers Rader, N. & Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture* 10, 202–221.
- Vouloumanos, A. & Werker, J.F. (2007). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science* 7, 270–276.
- Zacks, J.M. & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin* 127, 3–21.
- Zacks, J.M. & Swallow, K.M. (2007). Event segmentation. *Current Directions in Psychological Science* 16, 80–84.
- Zukow-Goldring, P. (1996). Sensitive caregiving fosters the comprehension of speech: When gestures speak louder than words. *Early Development and Parenting* 5, 195–211.
- Zukow-Goldring, P. (2006). Assisted imitation: Affordances, effectivities and the mirror system in early language development. In M.A. Arbib (ed.), *From action to*

*language via the mirror neuron system* (469–500). New York: Cambridge University Press.

## **Résumé**

*Les travaux antérieurs montrent que la synchronicité entre des événements émanant de deux modalités différentes constitue un concept-clé dans l'apprentissage social précoce. Notre étude pilote menée sur 14 dyades mère-enfant est la première à soutenir l'idée que la synchronicité entre action et langage en tant que forme de comportement responsable chez les mères est liée à l'acquisition ultérieure du langage de leurs enfants. Nous avons procédé à un codage détaillé du comportement multimodal au sein de la dyade dans le cadre de l'activité quotidienne du changement de couches lorsque les enfants étaient âgés de trois et de six mois. Lorsque les enfants ont atteint 24 mois, leurs mères ont renseigné des questionnaires sur le langage. Ces données ont été mises en relation avec les mesures dyadiques effectuées durant les interactions précoces. Nous proposons un modèle 'd'inversion des rôles' selon lequel il est important pour un enfant de trois mois d'être exposé plus longtemps à des inputs multimodaux, tandis que pour les enfants de six mois, la mère devrait répondre à l'attention de l'enfant et lui fournir des inputs multimodaux quand son enfant la regarde.*

Subj. #	Maternal synchrony % over interaction time in sec		Infants' attention to synchrony % over total infant gaze in sec		Attended synchrony % over total synchrony in sec		ELFRA-2 prod.	ELFRA- 2 verbs %
	3 M	6 M	3 M	6 M	3 M	6 M	24 M	24 M
01	43.88	44.25	56.84	59.49	54.45	75.51	200	14.50
02	21.60	26.04	43.03	32.63	61.56	34	174	14.37
03	13.93	17.94	13.05	33.55	19.92	24.12	66	4.55
04	30.05	32.94	33.03	37.26	83.13	48.09	45	6.67
05	43.64	41.93	47.95	53.98	52.30	58.28	191	16.75
06	42.37	44.02	44.52	58.28	63.64	38.33	160	14.38
07	27.41	11.66	29.97	28.74	71.63	29.33	99	6.06
08	21.67	15.59	26.39	28.51	69.84	46.49	215	18.14
09	22.53	20.31	25.17	29.83	97.51	23.47	27	0
10	43.74	34.44	56.55	43.56	43.77	59.29	81	8.64
11	30.76	30.66	31.08	32.98	94.94	33.97	136	11.03
12	33.02	24.08	40.31	29.74	77.94	46.47	210	18.57
13	20.35	0	19.89	0	77.98	0	16	6.25
14	36.50	30.22	41.77	41.48	73.21	45.70	157	17.20
MIN	13.93	0	13.06	0	19.92	0	16	0
MAX	43.87	44.25	56.83	59.49	97.51	75.51	215	19
MEAN (SD)	30.82 (10.1)	26.72 (12.9)	36.40 (13.1)	36.43 (15.1)	67.27 (20.5)	40.22 (18.5)	126.93 (70.2)	11 (5.87)

**Table 1: The dependent measures**

		Maternal synchrony		Infant's attention to synchrony		Attended synchrony		Maternal talkativeness		ELFR A-2 prod	ELFRA-2 verbs %
		3 M	6 M	3 M	6 M	3 M	6 M	3 M	6 M	24 M	24 M
Maternal synchrony	3M	1	.815** *	.888** *	.760**	-.078	.753**	.875** *	.716**	.420	.460
	6M		1	.783**	.941** *	-.190	.792**	.829** *	.921** *	.463	.418
Infant's attention to synchrony	3M			1	.722**	-.163	.808** *	.721**	.680**	.525*	.530*
	6M				1	-.352	.802**	.690**	.871** *	.502	.389
Attended synchrony	3M					1	-.260	.055	-.025	-.129	-.089
	6 M						1	.593*	.794**	.619*	.552*
Maternal talkativeness	3 M							1	.795**	.209	.320
	6 M								1	.433	.376
ELFRA-2 prod	24 M									1	.926***
ELFRA-2 verbs %	24 M										1

**Table 2: Correlations among the variables: the p-value is indicated**

by \*\*\* $p < .001$ ; \*\* $p < .01$ ; \* $p < .05$



**Figure 1: Camera setup**



**Figure 2: Example of how mothers synchronize their language with bodily actions (maternal synchrony):** In (a) there is a pause and the mother is in a neutral starting position. In (b) the mother takes an exaggerated inbreathe, during which she activates the face, moves her head back and takes the infant's legs lifting them up. In (c) while uttering the syllable "*ke*" she adds a further modality namely a torso movement, while at the same time moving the infant's legs outwards. Finally, in (d) while finishing the utterance, she returns to her starting position.

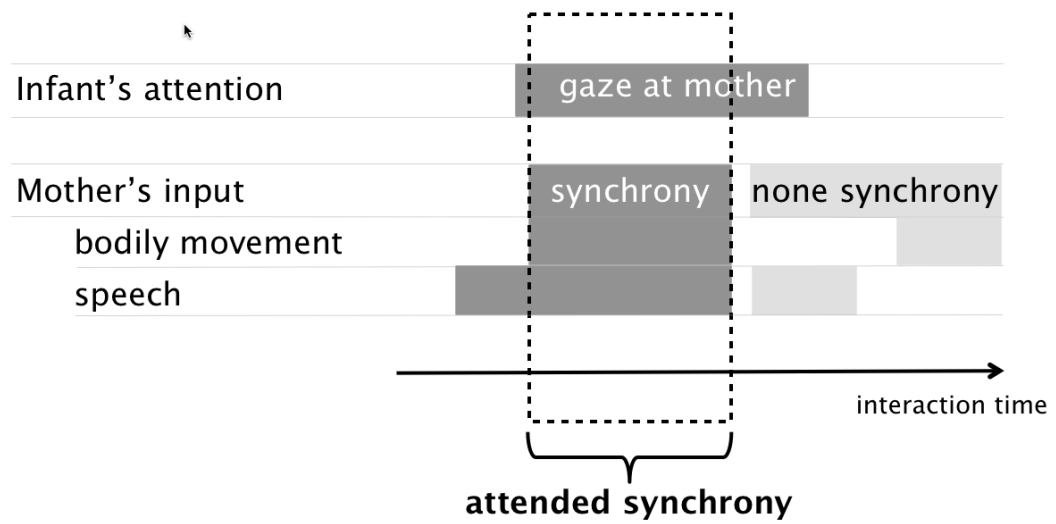


Figure 3: Schematic illustration of measures

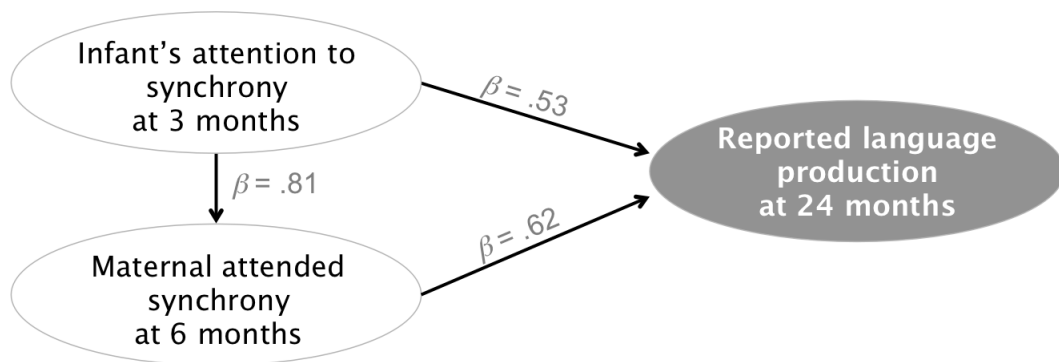


Figure 4: 'Switching the roles'-model